Journal of Computational and Applied Mathematics 9 (1983) 201–204
North-Holland

A perturbation analysis of the intrinsic conditioning of an approximate null vector computed with a SVD *

David W. KAMMLER
Department of Mathematics, Southern Illinois University, Carbondale, IL 62901, USA

Received 7 May 1982
Revised 21 February 1983

Abstract: Let A be an $m \times n$ real matrix with singular values $\sigma_1 \geq \cdots \geq \sigma_{n-1} > \sigma_n \geq 0$. In cases where $\sigma_n = 0$, the corresponding right singular vector v_n is a natural choice to use for an approximate null vector of A. Using an elementary perturbation analysis, we show that $\kappa = \sigma_1/(\sigma_{n-1} - \sigma_n)$ provides a quantitative measure of the intrinsic conditioning of the computation of v_n from A.

Keywords: Condition number, null vector, SVD.

AMS (MOS) Subject Classification: Primary 65F20, 15A12.

1. Introduction

Let A be a real $m \times n$ matrix, $m \geq n - 1$, let $u_1, \ldots, u_m$ and $v_1, \ldots, v_n$ be orthonormal bases for $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, and let A have the singular value decomposition

$$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_n u_n v_n^T \tag{1}$$

where $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$, cf. [1,2]. We thus have

$$Av_k = \sigma_k u_k, \quad k = 1, \ldots, n, \tag{2}$$

$$A^T A v_k = \sigma_k^2 v_k, \quad k = 1, \ldots, n. \tag{3}$$

By using (3) we see that

$$\min\{\|Ax\|_2 : x \in \mathbb{R}^n \text{ and } \|x\|_2 = 1\} = \sigma_n \tag{4}$$

with the minimum occuring when x is a normalized right singular vector of A corresponding to the singular value σ_n.

When $\sigma_n = 0$, v_n is an approximate null vector of A which is optimal in the sense that this choice yields the minimum in (4). For this reason, the SVD is sometimes used as a tool for computing an approximate null vector for a given matrix A, especially in cases where this null vector gives rise to certain physically significant parameters [3].

In this paper we analyze the way in which the approximate null vector v_n is changed by small perturbations in A. In so doing we assume that $\sigma_{n-1} > \sigma_n$ in order to insure that v_n is uniquely determined (to within a scale factor). Within this context we will show that the condition number

$$\kappa = \sigma_1/(\sigma_{n-1} - \sigma_n) \tag{5}$$

* This research was supported in part by the Rome Air Development Center, Griffiss AFB, NY, under contract F 30602-78-C-0148.

provides a quantitative measure of the intrinsic (i.e. algorithm independent) local conditioning of this calculation of v_n from A.

2. A perturbation analysis

Using an elementary perturbation analysis we develop the following first-order bound.

Theorem 2.1. *Let A, E be real $m \times n$ matrices with $m \geq n-1$. For each ϵ in some neighborhood of $\epsilon = 0$ let the matrix $A + \epsilon E$ have the singular values*

$$\sigma_1(\epsilon) \geq \cdots \geq \sigma_{n-1}(\epsilon) > \sigma_n(\epsilon) \geq 0$$

and let $v_n(\epsilon)$ be the right singular vector corresponding to $\sigma_n(\epsilon)$. Then

$$\|v_n(\epsilon) - v_n(0)\|_2 \leq \frac{\|E\|_2 \epsilon}{\sigma_{n-1} - \sigma_n} + O(\epsilon^2) \quad \text{as} \quad \epsilon \to 0^+. \tag{6}$$

(Here $\|E\|_2$ denotes the spectral norm of the matrix E.)

Proof. Although (6) can be obtained within the more general context of [5,6], we shall present an alternative direct argument. We write

$$v_n(\epsilon) = v_n + (\alpha_{11}v_1 + \cdots + \alpha_{n1}v_n)\epsilon + (\alpha_{12}v_1 + \cdots + \alpha_{n2}v_n)\epsilon^2 + \cdots, \tag{7}$$

$$\sigma_n(\epsilon) = \sigma_n + \sigma_{n1}\epsilon + \sigma_{n2}\epsilon^2 + \cdots$$

where $\sigma_1 \geq \cdots \geq \sigma_{n-1} > \sigma_n \geq 0$ are the singular values and $v_1, \ldots, v_n$ the corresponding orthonormal right singular vectors of the unperturbed matrix A, cf. [4, Theorem 7.7.1]. In analogy with (3) we write

$$(A + \epsilon E)^{\mathrm{T}}(A + \epsilon E)\left[v_n + (\alpha_{11}v_1 + \cdots + \alpha_{n1}v_n)\epsilon + \cdots\right] =$$

$$= (\sigma_n + \sigma_{n1}\epsilon + \cdots)^2\left[v_n + (\alpha_{11}v_1 + \cdots + \alpha_{n1}v_n)\epsilon + \cdots\right],$$

and after using (3) to simplify the equation which results when we equate the coefficients of ϵ on the two sides of this equation, we obtain the first-order identity

$$(A^{\mathrm{T}}E + E^{\mathrm{T}}A)v_n = -(\sigma_1^2 - \sigma_n^2)\alpha_{11}v_1 - \cdots - (\sigma_{n-1}^2 - \sigma_n^2)\alpha_{n-1,1}v_{n-1} + 2\sigma_n\sigma_{n1}v_n.$$

Since $v_1, \ldots, v_n$ are orthonormal, we immediately obtain the expressions

$$\alpha_{k1} = -\frac{v_k^{\mathrm{T}}(A^{\mathrm{T}}E + E^{\mathrm{T}}A)v_n}{\sigma_k^2 - \sigma_n^2}, \quad k = 1, \ldots, n-1,$$

which we may further simplify to

$$\alpha_{k1} = -\frac{\sigma_k u_k^{\mathrm{T}}Ev_n + \sigma_n u_n^{\mathrm{T}}Ev_k}{\sigma_k^2 - \sigma_n^2}, \quad k = 1, \ldots, n-1$$

by using (2). Assuming (7) is normalized, i.e.

$$\|\alpha_{11}\epsilon v_1 + \cdots + \alpha_{n-1,1}\epsilon v_{n-1} + (1 + \alpha_{n1}\epsilon)v_n + O(\epsilon^2)\|_2^2 = 1 + 2\alpha_{n1}\epsilon + O(\epsilon^2) = 1,$$

we see that $\alpha_{n1} = 0$. Our perturbation analysis thus leads to the expression

$$v_n(\epsilon) = v_n - \epsilon \sum_{k=1}^{n-1}\left[\frac{\sigma_k u_k^{\mathrm{T}}Ev_n + \sigma_n u_n^{\mathrm{T}}Ev_k}{\sigma_k^2 - \sigma_n^2}\right]v_k + O(\epsilon^2)$$

with

$$\|\boldsymbol{v}_n(\epsilon)-\boldsymbol{v}_n\|_2^2=\epsilon^2\cdot\sum_{k=1}^{n-1}\left[\frac{\sigma_k\beta_{kn}+\sigma_n\beta_{nk}}{\sigma_k^2-\sigma_n^2}\right]^2+O(\epsilon^3) \tag{8}$$

where for notational convenience we set

$$\beta_{kn}=\boldsymbol{u}_k^{\mathrm{T}}E\boldsymbol{v}_n,\qquad \beta_{nk}=\boldsymbol{u}_n^{\mathrm{T}}E\boldsymbol{v}_k,\quad k=1,\dots,n-1.$$

We must now show that the precise but cumbersome relation (8) can be replaced by the simple bound (6).

As a first step we use Bessel's inequality to write

$$\sum_{k=1}^{n-1}\beta_{kn}^2=\sum_{k=1}^{n-1}|\boldsymbol{u}_k^{\mathrm{T}}(E\boldsymbol{v}_n)|^2\leqslant\|E\boldsymbol{v}_n\|_2^2\leqslant\|E\|_2^2,$$

$$\sum_{k=1}^{n-1}\beta_{nk}^2=\sum_{k=1}^{n-1}|(E^{\mathrm{T}}\boldsymbol{u}_n)^{\mathrm{T}}\boldsymbol{v}_k|^2\leqslant\|E^{\mathrm{T}}\boldsymbol{u}_n\|_2^2\leqslant\|E\|_2^2. \tag{9}$$

We then use (8)–(9), Cauchy's inequality, and the bounds

$$\frac{\sigma_k}{\sigma_k^2-\sigma_n^2}\leqslant\frac{\sigma_{n-1}}{\sigma_{n-1}^2-\sigma_n^2},\quad k=1,\dots,n-1,$$

$$\frac{\sigma_n}{\sigma_k^2-\sigma_n^2}\leqslant\frac{\sigma_n}{\sigma_{n-1}^2-\sigma_n^2},\quad k=1,\dots,n-1$$

to write

$$\begin{aligned}\epsilon^{-2}\|\boldsymbol{v}_n(\epsilon)-\boldsymbol{v}_n\|_2^2&=\sum_{k=1}^{n-1}\left\{\left[\frac{\sigma_k}{\sigma_k^2-\sigma_n^2}\right]^2\beta_{kn}^2+2\left[\frac{\sigma_k}{\sigma_k^2-\sigma_n^2}\right]\left[\frac{\sigma_n}{\sigma_k^2-\sigma_n^2}\right]\beta_{kn}\beta_{nk}+\left[\frac{\sigma_n}{\sigma_k^2-\sigma_n^2}\right]^2\beta_{nk}^2\right\}+O(\epsilon)\\
&\leqslant\frac{\sigma_{n-1}^2}{(\sigma_{n-1}^2-\sigma_n^2)^2}\sum_{k=1}^{n-1}\beta_{kn}^2+\frac{2\sigma_{n-1}\sigma_n}{(\sigma_{n-1}^2-\sigma_n^2)^2}\cdot\left[\sum_{k=1}^{n-1}\beta_{kn}^2\sum_{k=1}^{n-1}\beta_{nk}^2\right]^{1/2}\\
&\quad+\frac{\sigma_n^2}{(\sigma_{n-1}^2-\sigma_n^2)^2}\sum_{k=1}^{n-1}\beta_{nk}^2+O(\epsilon)\\
&\leqslant\frac{\|E\|_2^2(\sigma_{n-1}+\sigma_n)^2}{(\sigma_{n-1}^2-\sigma_n^2)^2}+O(\epsilon)\\
&=\frac{\|E\|_2^2}{(\sigma_{n-1}-\sigma_n)^2}+O(\epsilon),\end{aligned}$$

thus obtaining (6). □

Note (1) When $n\geqslant 2$ and the perturbation is chosen so that

$$E\boldsymbol{v}_k=\boldsymbol{0}\quad\text{for } k<n-1,$$

$$E\boldsymbol{v}_{n-1}=\boldsymbol{u}_n,\qquad E\boldsymbol{v}_n=\boldsymbol{u}_{n-1},$$

we verify that

$$\beta_{kn}=\beta_{nk}=0\quad\text{for } k<n-1,$$

$$\beta_{n-1,n}=\beta_{n,n-1}=1=\|E\|_2,$$

and then use (8) to conclude that

$$\|\boldsymbol{v}_n(\epsilon)-\boldsymbol{v}_n\|_2=\frac{\|E\|_2\epsilon}{\sigma_{n-1}-\sigma_n}+O(\epsilon^2)\quad\text{as}\quad\epsilon\to 0^+.$$

In this way we see that the bound (6) is sharp.

(2) In the degenerate case where $\sigma_1 \geqslant \cdots \geqslant \sigma_{n-p} > \sigma_{n-p+1} = \cdots = \sigma_n \geqslant 0$ for some p with $1 < p < n$, a slight extension of the above analysis (based on [4, Theorem 7.10.1]) can be used to show that for a suitable choice of $v_{n-p+1}, \ldots, v_n$ we have

$$\|v_k(\epsilon) - v_k(0)\|_2 \leqslant \frac{\|E\|_2 \epsilon}{\sigma_{n-p} - \sigma_n} + O(\epsilon^2) \quad \text{as} \quad \epsilon \to 0^+ \tag{10}$$

for each $k = n-p+1, \ldots, n$.

(3) In some cases one is interested in the angle $\phi(\epsilon) \geqslant 0$ between $v_n(\epsilon)$ and $v_n(0)$. In view of (6) we have

$$\begin{aligned} \sin \phi(\epsilon) &= \|v_n(\epsilon) - v_n(0)\|_2 / \|v_n(0)\| + O(\epsilon^2) \\ &\leqslant \frac{\|E\|_2 \epsilon}{\sigma_{n-1} - \sigma_n} + O(\epsilon^2) \quad \text{as} \quad \epsilon \to 0^+. \end{aligned}$$

More generally, in the p-fold degenerate case of (10) the angle $\phi(\epsilon)$ between the subspaces spanned by $v_k(\epsilon)$, $k = n-p+1, \ldots, n$, and $v_k(0)$, $k = n-p+1, \ldots, n$, satisfies

$$\sin \phi(\epsilon) \leqslant \frac{\sqrt{p}\,\|E\|_2 \epsilon}{\sigma_{n-p} - \sigma_n} + O(\epsilon^2) \quad \text{as} \quad \epsilon \to 0^+,$$

in conformity with the generalized sin θ theorem of [6, p. 102].

3. The local condition number

When we calculate v_n from A (e.g. using the software in [1]) we must certainly expect to deal with perturbations of size $\|E\|_2 \epsilon$ in A where σ is the unit roundoff of the computer we are using and where

$$\|E\|_2 \simeq \|A\|_2 = \sigma_1.$$

In view of (6), we must then expect to encounter an error of approximate size $\kappa\epsilon$ in our approximate null vector v_n, with κ being given by (5). If A actually has a null vector, i.e. $\sigma_n = 0$, and if $\sigma_1 = \cdots = \sigma_{n-1} > 0$, then we obtain the smallest possible $\kappa = 1$. In practice, we often encounter situations where $\sigma_1, \sigma_2, \ldots$ decrease fairly rapidly so that $\kappa \gg 1$ and v_n is poorly determined. In such cases the size of κ gives us a useful quantitative measure of the intrinsic local conditioning of the null vector calculation and helps us to assess the accuracy of a computed approximation to v_n.

References

[1] J.J. Dongarra, C.B. Moler, J.R. Bunch and G.W. Stewart, *LINPACK User's Guide* (SIAM, Philadelphia, PA, 1979).

[2] G.E. Forsythe, M.A. Malcolm and C.B. Moler, *Computer Methods for Mathematical Computation* (Prentice-Hall, Englewood Cliffs, NJ, 1977).

[3] T.L. Henderson, Geometric methods for determining system poles from transient response, *IEEE Trans. Acoust. Speech Signal Process.* **29** (1981) 982–988.

[4] P. Lancaster, *Theory of Matrices* (Academic Press, New York, 1969).

[5] G.W. Stewart, Error and perturbation bounds, SIAM Rev. **15** (1973) 727–764.

[6] P. Wedin, Perturbation bounds in connection with singular value decomposition, *BIT* **12** (1972) 99–111.